

# **SUMMARY OF THE EPA-SPONSORED TERRESTRIAL PEER INPUT WORKSHOP**

**By Dr. Andrew H. Macdonald**

**July 11, 1999**

## **INTRODUCTION**

On June 22-24, 1999, the Environmental Protection Agency (EPA) sponsored two public workshops as part of its initiative to revise the ecological assessment process for pesticides. The goal of this initiative is to identify, develop, and validate tools and methodologies to conduct probabilistic assessments and improve risk characterization. These assessments would be used to address the impacts of pesticides on nontarget organisms.

A key component of this initiative is the Ecological Committee on FIFRA Risk Assessment Methods (ECOFRAM), which is composed of experts drawn from a variety of stakeholders, including government agencies, academia, environmental groups, industry, and others. They are divided into the Aquatic and Terrestrial Workgroups and have been working over the last few months to develop recommendations for EPA's consideration in revising the assessment process. They have discussed potential probabilistic tools and methodologies and have summarized their findings in draft reports.

The purpose of the workshops was to provide peer input on these draft reports. Specifically, the workshops would provide EPA with external scientific review, comment, and discussion of the draft reports, which EPA would factor into decisions regarding the implementation of ECOFRAM recommendations. In addition, it would provide the ECOFRAM members with comments to help them finalize the reports.

This report provides a summary of the workshop held to address the terrestrial draft ECOFRAM report.

## **PANEL PRESENTATIONS**

**Ms. Denise Keehner, Acting Director, Environmental Fate and Effects Division, Office of Pesticide Programs, U.S. EPA**

In her opening remarks, Ms. Keehner noted that in May 1996 EPA asked the FIFRA Scientific Advisory Panel to review and comment on the Office of Pesticide Program's ecological risk assessment methods and procedures. While recognizing and generally reaffirming the utility of the assessment process currently used, the Panel encouraged the Agency develop the tools and methodologies necessary to conduct probabilistic assessments.

In response to the recommendations of the SAP, OPP began a new initiative to develop and validate tools and methodologies to conduct probabilistic assessments to address terrestrial and aquatic risk. The purpose of this initiative is to strengthen the core elements of the ecological

assessment process by developing and validating probabilistic assessment tools and methodologies. These methodologies are intended for use by OPP for evaluating effects of pesticides to terrestrial and aquatic species.

Ms. Keehner acknowledged the hard work and efforts of ECOFRAM, which have resulted in the completion of their draft aquatic and terrestrial reports. The purpose of this workshop is for the Peer Input Panel to review the Terrestrial Workgroup's report. The Panel was specifically asked to consider these five questions:

- \* Is the report scientifically sound?
- \* Did the ECOFRAM working group fulfill its goals?
- \* What are the limitations to this approach?
- \* What areas need to be strengthened?
- \* How should this process be integrated in the overall risk management process?

**Ms. Sandra L. Bird, Environmental Engineer, U.S. EPA/ORD, Ecological Research Division, Athens, Georgia**

Although Ms. Bird found the report too detailed, she thought that it was still a step forward for the risk assessment process. She felt that the report's focus should be broadened generally and that there should be a greater attempt to link it to the Aquatic ECOFRAM study. The risk assessment and decisionmaking structure of the two reports needs to be brought closer together.

In considering whether ECOFRAM had fulfilled its charge, she noted that a variety of other exposure pathways -- including earthworms, mammals, amphibians, and reptiles -- were not included in the analysis; that there should be a discussion of why certain species are considered "threshold" (endpoint) species; that there was not really any protocol for selecting the "endpoints"; and that it lacked a good framework for determining avian exposure and toxicity patterns.

The report did not consider how to define threshold effects at the population level -- at both regional and local scales -- or on the basis of a risk benefit analysis. How the "threshold of acceptability" is defined is a key part of probabilistic risk assessments. Probabilistic assessments should be broadened to include other species, indirect effects, spatial effects, and habitat effects.

She thought that there were too many experts in the field of avian toxicity on the committee and that the report should have a broader ecological perspective. Case studies involving population effects and spatial factors, such as the range of a species and habitat destruction, were not given enough attention in the report.

In conclusion, risk assessments should focus on population effects, not only the death of individuals. Risk assessments should not ignore indirect, relative, and situational factors. She suggested, for example, that when the habitat of a species is already threatened, pesticide mitigation may not be a sound investment.

**Dr. Timothy Barry, Office of Economy and Environment, Climate and Policy Assessment Division, U.S. EPA**

Dr. Barry's discussion focused on the characteristics of a good Monte Carlo simulation. In evaluating the report's scientific soundness he looked for several things: sound methods, reproducible results, and transparency. When asked if the report was scientifically sound, he said "maybe." Monte Carlo provides a generally sound framework for probability analysis, but the "devil is in the details," he noted.

One must first determine if the model has a clear purpose and whether the endpoints have been clearly defined. Have the populations at risk been fully identified? Have the methods, models, data, and assumptions, including sensitivity analysis, been well documented?

He described several important and contentious Monte Carlo issues, including:

- \* selection of appropriate exposure models
- \* characterization of model data
- \* parameter correlation
- \* uncertainty characterization
- \* model validation

How one selects a dose-response model is very important. Data fitting must be accompanied by good statistical analysis and should include a careful assessment of the deviation at the lower boundaries of the distribution. The use and impact of expert judgment in situations where there is insufficient information needs to be carefully discussed. When surrogate data are used, their use must be accompanied by an explanation.

The analysis should be limited to critical exposure and effects pathways, and sensitivity analysis should be used to guide decisions about important model structures, input assumptions, and parameters. There should be a thorough discussion of how input variables and effect are related.

Variability should be separated from uncertainty, and when uncertainties cannot be quantified there should be at least a qualitative review. This issue was not adequately discussed in the ECOFRAM report according to Dr. Barry. He recommended using the "population risk model" of Bogen and Spear (1987) to generate "probability density functions" (PDFs). More attention should be given to how and when uncertainty is determined for parameters other than exposure and risk models. Bounding techniques should be used in situations where there are insufficient data.

Without real-world case studies, it is hard to fully evaluate the soundness of the risk probability techniques described in the report. Ultimately, the scientific credibility of the report will depend on how the models are applied, how the process is verified, and whether other direct and indirect effects are included in the model.

**Dr. George P. Cobb, Institute for Environmental and Human Health, Texas Tech University and the Texas Tech Health Science Center**

Pesticide exposure routes and their toxic effects are covered fairly well in the report. However, the degradation and dissipation rates (fate) and interspecies toxicity effects of pesticides must be considered as well. Moreover, probabilistic risk assessment methods are already part of many existing fate and transport models.

The report did not consider multiple exposure scenarios in several crops, and nondermal (e.g., soil ingestion) exposure routes, the impact of pesticides on other invertebrates and the synergistic effects between pesticides. Furthermore, to be valid, dose-response tests should be conducted on more than one species of concern. A good model should be able to account for all invertebrate prey species (e.g., arachnids), the effect of random pest ingestion, and rapidly changing exposure levels, which often vary with respect to time and geography following the application of a pesticide.

In conclusion, holistic models need lots of real-world data.

**Peter Edwards, Zeneca Agrochemicals**

The ECOFRAM report represents a significant step forward in the way we conduct terrestrial risk assessment. However, the assessment focus should be broadened beyond oral effects and birds and needs to address the impact of pesticides through other exposure routes, in particular that involving the skin and lungs. Indirect effects should be analyzed too.

Higher level probabilistic risk assessments will be limited by the amount of data that will have to be generated for such assessments. Deterministic models do work in some cases, he believes, because he does see healthy birds. Yet, the "levels of refinement" process will enable us to match the data generated by the assessment to the level of knowledge needed. There needs to be a concerted effort to gather "generic" data distributions. The "devil is in the distributions," he suggested.

The "threshold of sensitivity" must be well defined -- it should encompass other perturbations affecting agricultural landscapes. Value judgments will play an important role in the decisionmaking process. In general, the report is not well organized. Because it is so overloaded with details, key points are hard to recognize; it is hard to know if and where old tests are still valid. The report also suggests -- perhaps incorrectly -- that the amount of uncertainty is so great that it cannot be handled scientifically.

Perhaps pesticides are not as important as landscape management and agricultural intensification.

**Dr. Chris Grue, Leader, Washington Cooperative Fish and Wildlife Research Unit,  
University of Washington**

Despite the time limitations of the committee and the somewhat reduced scope of the analysis, Dr. Grue thought the ECOFRAM report was a "state of the art" review of probabilistic risk assessment.

Still, he is frustrated by the lack of real progress made by EPA and industry over the last decade. Many of the recommendations previously made the Avian Effects Dialogue Group (ADEG), such as collection of better field data and sharing of data on avian populations, still have not been implemented. He believes that the shift toward using laboratory toxicity data, in order to meet registration requirements, was harmful.

Perhaps the most important part of the report is the identification of data ("data gaps") needed to reduce the "uncertainty" in risk predictions. Collecting this information will require EPA and industry to work more closely together. He does not believe that probabilistic risk assessments can be conducted using existing data alone, as the report tends to suggest. Focused field studies are needed urgently in order to improve both exposure and effects predictions. Indeed, the report should have contained several complete case studies.

Exposure and effects predictions can be improved by:

- \* including chemical additives that are mixed with pesticides -- many are toxic themselves
- \* including the ingestion of contaminated soil by invertebrates and vertebrates
- \* considering indirect ecological effects

If there are better exposure data and dose response curves for selected species, a level 3 risk analysis is possible. Determining the acceptable risk is more complex -- it will depend on the resource that is threatened. It may be necessary to be more conservative when the species at risk are "threatened and endangered species." He favors conditional registration of pesticides, as exists today in the UK.

In conclusion:

- \* We need to conduct several real-world case studies.
- \* Indirect ecological effects are too important to be ignored.
- \* We need to consider the Western European approach to risk assessments, in which risk is less prejudiced by the "value" of individual species and more concerned with ecosystem effects generally.

**Dr. Michael L. Lavine, Institute for Statistics and Decision Science, Duke University**

Dr. Lavine was concerned with how the ECOFRAM report accounts for the inherent uncertainty in toxicity parameters such as the LD50. He noted that a single toxicity study, or even a collection of such studies, cannot give us the exact LD50 or the slope of a dose-response

curve, since they reveal only what conclusion is most likely given the data that are currently available. The uncertainty in extrapolated data can be reduced by using probability distributions.

Instead of using point estimates in data-poor cases, he recommends using probability distributions that assign more weight to data-rich studies and less weight to those where data are lacking. In other words, we should not bias our conclusions regarding uncertainty by assuming that we understand what the distributions are beforehand. Instead, a range of distributions -- perhaps a conservative range -- should be considered and their extrapolation value compared using a sensitivity (or Bayesian average) analysis.

Moreover, he believes that the actual accuracy of all probabilistic assessments (e.g., of risk, dose, and toxicity) should be estimated as well, and the range of possible risks given. Prediction intervals and Bayesian methods are useful tools for estimating the accuracy of such information.

A good risk assessment should be able to handle parameter uncertainty, model uncertainty, and the uncertainty resulting from data extrapolation. Thus it is important not to rely on one model -- only probit for example -- to generate dose-response curves and toxicity endpoints. Instead, the toxicity predictions from several models -- probit, logic and others -- should be compared.

In conclusion, there are many other ways besides Monte Carlo simulations to integrate a range of data points. For example, the variability in toxicity measurements might be analyzed more effectively using a hierarchical model.

**Dr. Robert Luttik, Center for Substances and Risk Assessment, National Institute of Public Health and the Environment, The Netherlands**

Dr. Luttik thought that the ECOFRAM report provided a good framework for conducting probabilistic risk assessments. However, the report needs to give as much attention to reptiles and mammals as it does to birds. In order to conduct higher level risk assessments, a lot more applied research on the effects of pesticides on terrestrial invertebrates, in particular, needs to be done.

Risk assessment models need to consider the secondary impact of pesticides on birds of prey, small insect-eating passerine, and grain-eating birds (and smaller mammals). He discussed a study in which the assimilation of pesticides was correlated to metabolic rate, caloric intake, food assimilation efficiency, species sensitivity, and pollutant uptake efficiency. Such studies can be used to build a probabilistic soil-contamination model involving predator species like those listed above and others like grass-eating geese, which must eat a lot of food to survive.

In calculating exposure levels, he believes it might be better to use the "pesticide weight (in mg) per kilojoule per day" rather than "pesticide weight per body weight per day."

In regard to exposure, it appears that some birds eat more contaminated food than pesticide-free food; or that at least it varies over time. He is not convinced that each species of bird prefers a particular grit size. In general, Dr. Luttik recognizes the need to conduct a great deal more research on exposure rates and pathways and species (endpoints), especially for higher level risk assessments. Initial risk mitigation decisions can probably be made using tier (level) 2 risk assessments, as described in the ECOFRAM report.

**Dr. Dwayne Moore, The Cadmus Group, Ottawa, Ontario**

Dr. Moore thought the report was generally scientifically sound, but that the range of methods used to predict effects and risks was too limited.

A Monte Carlo simulation is not necessarily the best way to characterize uncertainty; Bayesian probability theory (see his remarks in Appendix) and expert knowledge are also important tools. Indeed, many sources of uncertainty are nonprobabilistic. He also questioned the validity of using conservative assumptions in a probabilistic risk assessment when data are missing, since that is more of a policy (risk management) decision.

He felt that the problem being addressed in the ECOFRAM report was well formulated but that concepts like "assessment endpoint" were confusing (see risk scenario terms of Kaplan and Garick cited in his paper). The risk management questions in Chapter 2 are very good and should have been answered in the report. The section on exposure is excellent, but it would be improved by a discussion about how to select appropriate data and whether models (like PRZM) have been field tested; the uncertainty regarding model structure was not discussed.

Large fate and transport models, with many variables, like some of the water quality models, have performed poorly, he said. Expert knowledge, which can improve the design of models, is often overlooked.

The "effects" section was the weakest: it did not adequately consider nonlethal effects; it considered only birds and the probit model; and it did not discuss how to obtain toxicity data for species such as mammals. More comprehensive "effects" data sets need to be acquired or are already available, like that currently available for high-use pesticides such as the triazines. The Generalized Linear Model (GLIM) is a useful tool for calculating dose-response relationships; dose-response tests for organisms other than birds need to be developed and the data collected.

Finally, it is not clear how risk managers will use probabilistic-based assessments to make decisions or how the results of such an analysis will be presented to other important audiences, including the public at large.

**Dr. Edward W. Odenkirchen, Biologist, Environmental Fate and Effects Division,  
Office of Pesticide Programs, U.S. EPA**

Dr. Odenkirchen confined his review to exposure assessments and the "level of refinement" process. He noted that the report focused on oral exposure (ingestion) routes at the

expense of other pathways, including multiple exposure routes. Indirect effects were not considered yet the models used in the early stages of a risk assessment are still very complex.

The exposure models developed in the ECOFRAM report need to be validated using empirical data and case studies involving many chemicals and multiple exposure and effects pathways. Unfortunately, EPA has not established a good framework for collecting these data; moreover, a lot of the existing data are too site specific. Indeed, the report did not address either indirect or large-scale ecological effects, as outlined in the Background document. Furthermore, the framework established for refining exposure assessments is too flexible, not structured enough, and should be modified.

Risk predictions, especially at higher refinement levels, are hindered by the limited amount of exposure data available. At present there are only enough oral exposure data for level 1 and 2 risk assessments. The uncertainty caused by data extrapolation must be reduced and the uncertainty related to model structure must be evaluated.

Furthermore, both the "threshold of acceptability" and the cost of mitigation itself must be considered when making a risk management decision. Indeed, he believes that mitigation decisions will determine just how many exposure data are required. For example, in cases where mitigation involves changing the timing of an application, the risk assessment should focus on impacts to key species.

A number of issues should be studied in more detail, including:

- \* Are level 1 exposure assessments always conservative?
- \* When is a more refined assessment appropriate?
- \* Who should make the decision to reduce the uncertainty further?
- \* How can we improve the quality of exposure parameters?
- \* Are the assumptions regarding the use of the avoidance factor correct?
- \* Are the assumptions about diet preferences and food intake for birds accurate?
- \* How does one estimate the combined impact of several exposure pathways? Can we rely upon only one pathway -- such as ingestion -- to characterize the total risk?

In conclusion, Dr. Odenkirchen favors a "tier" structure, rather than the flexible system levels recommended in the report, to guide refinements in probabilistic risk assessments and the associated mitigation process.

**Dr. Glenn Suter, Ecologist, National Center For Environmental Assessment, Cincinnati, Ohio**

Dr. Suter thought that the ECOFRAM report highlighted a number of potentially useful methods for improving risk assessment decisions. But he is concerned about how these methods will be used by regulators to improve risk management decisions: Should risk predictions be given to risk managers at each stage of review? Should we be comparing related pesticides when making decisions?



He noted that it is very important to define appropriate ecological "endpoints" before deciding how to determine probabilistic effects. Should we be concerned primarily about the loss of individual species, or the probability of levels of effects? Threatened and endangered species have not been seen as distinct endpoints. When are population effects important? Are focal species, or the class of species they represent, the real variable of concern? Should endpoints be considered "thresholds," or continuous variables?

The report considered terrestrial birds as the primary organism of concern in the models. It did not consider a number of other important species and values, including piscivorous birds, mammals, reptiles, amphibians, plants, and ecosystem processes.

Many important routes of exposure -- including grooming and preening -- were not considered. It is very important to say why certain exposure and effects pathways are not that important. Additional research needs to be done to determine how "indirect" ecological effects such as the loss of seeds and insects in agricultural fields affect other species. When necessary, pesticide manufacturers should be required to provide data for ecologically significant issues. A combination of expert judgment and research can help provide answers to many of these "generic" effects questions.

Many other sources of information can be used to help develop better risk models, including mechanistic studies of related pesticides, QSAR, and toxicological studies on mammals. "Ecological studies alone will not provide all the answers," noted Dr. Suter.

In effect, the ECOFRAM authors said at some points "we are too uncertain to be uncertain." He believes "the more uncertain you are, the more you need uncertainty analysis." Therefore, the real value of using probabilistic (uncertainty) analysis will be to reveal the actual uncertainty in predictions of threshold (mitigation) level effects for individual pesticides. Models can also show where we need to improve risk assessments generally; it will be necessary to utilize tools other than Monte Carlo simulations and probabilistic methods.

The risk assessment process should not be driven by model design and complexity. Critical endpoints and data gaps should be identified before a particular method of analysis is selected. He envisions a flexible structure in which every critical threshold effect is considered at each stage of the assessment process. The tiers would be organized in terms of the quantity and cost of the data produced, rather than complexity of the assessment models. In instances where few data are available, the assessment may actually be more complex because of the requirement to use models in place of data to estimate risk.

In conclusion, it is important to first identify the overall goals of the assessment, then choose appropriate endpoints and data, then choose the models. The recent "ozone standards" decision by the DC Federal Court shows why a regulatory decision must be based on a very precise scientific understanding and assessment of the risk -- the fact that a decision was not reached in an "arbitrary and capricious" manner may no longer be enough to justify a particular risk management decision.

**Douglas J. Urban, Senior Scientist, Environmental Fate and Effects Division,  
Office of Pesticide Programs, U.S. EPA**

According to Dr. Urban, both risk managers and assessors will have to learn how to use complex mathematical tools like sensitivity analysis, probability density functions, and Monte Carlo simulations.

There is a critical need for additional data beyond those currently required for registration purposes. These new tools and methods will help reveal how much uncertainty actually exists in current risk predictions. He is not convinced that current deterministic-based assessments are overly conservative -- this assumption needs to be tested as soon as possible using probabilistic models.

The inhalation and dermal absorption of pesticides should be considered when determining both exposure levels (dose) and their effects. In situations where there are not enough empirical data, expert judgment will be very important. Higher level (tier) exposure assessments will require the collection of site-specific residue data and the expansion of databases for residues on insects, soils, vegetation, and soil invertebrates.

The avian reproduction test provides only a rough (screening) assessment of long-term effects. If refinement in level 2 is necessary, additional avian reproduction studies could be required using a dose-response design based on the most sensitive response observed in level 1 tests. New dose-response test methods need to be developed, including one for altricial birds, and the uncertainties surrounding the extrapolation of all laboratory results to other birds need to be understood. For acute risk level 2, the dose-response of four or more sensitive "endpoint" species should be determined.

The difference between lab and field studies is a great source of uncertainty in risk assessments, as is the failure to include indirect and sublethal effects of pesticides. Case studies are an urgent near-term need and should be developed to test these new risk prediction methods and to establish a consistent method of risk analysis.

Consistency and the reduction of uncertainty in risk assessments would be enhanced by a less flexible tier system, with similar levels of refinement (for both exposure and effects) -- rather than the more flexible approach advocated in the terrestrial ECOFRAM report. Establishing more rigorous "threshold" levels will be especially difficult given the political nature of the regulatory process and the cost of mitigation. Regulators will have to consider other factors as well, including proximity to sensitive habitats and endangered species. Scientists should begin by defining, if possible, lower ("minimum") threshold levels for populations of different species.

Remember that the results of a probabilistic risk assessment is not the final product, but must be weighed along with other nonquantitative information, such as incidents, in the risk characterization.

In conclusion, a major effort should be made to gather exposure and effects data

(especially pesticide residues) for use in these probabilistic models, to analyze several case studies (with the participation of risk managers), to study how pesticides affect other terrestrial species beside birds, and to look at exposure and effects pathways other than food.

## **PANEL DISCUSSION**

Questions posed to the panel were as follows:

1. Are the probabilistic methods presented in the ECOFRAM report effective for the endpoints that are defined in the report? What other methods and data could be used to conduct a probabilistic risk assessment of other endpoints?
2. Is the proposed level of refinement process practical and logical?
3. Are the data recommended by the terrestrial ECOFRAM report sufficient to develop probabilistic risk assessments?
4. Can reproductive, population, and ecosystem "endpoints" be assessed using probabilistic models?
5. Do you agree with the recommendations for future research and validation listed in Table 7.6.1?
6. How should case studies be conducted?

### **Question 1: ECOFRAM Endpoints**

The first panel discussion focused on several issues, ranging from what is an appropriate endpoint to the methods needed to conduct an effective probabilistic risk assessment at the individual, population, and community levels. Uncertainty and data extrapolation related to the selection of endpoints and the threshold effects were also discussed.

*Endpoints.* One panelist began the discussion by asking: "How far do you go in the risk assessment process before you make a decision regarding mitigation? Do you need to see adverse effects at the individual level? The ecosystem level?" Do you look at ecosystem level effects only during the risk characterization stage of the assessment?

Everyone recognized that a clear statement of the relevant endpoint species and effects (e.g., mortality) must occur early in the process. Terms like "valued ecological entity" are too vague.

The EPA has the regulatory responsibility to consider effects to individual species -- because of the Endangered Species Act -- and broader scale effects within agricultural environments and communities. In past registration cases -- such as that involving the pesticides diazinon and carbofuran used on golf courses and sod farms -- risk quotients were calculated using mortality data. Risk managers want to understand (indeed to be able to predict) larger scale ecological effects, in addition to widespread mortality to individuals.

*Population-level effects: too much uncertainty to be useful?* There was a great deal of

discussion about the value of conducting risk assessments involving population-level effects (etc.) in situations where there are only limited ecological data. Are predictions that use generic life histories valid, or useful? Some of the panelists thought that such "what if" stochastic models, although less precise because of the uncertainty present, would still provide regulators with useful estimates of risk.

The uncertainty present within "organism" level risk assessments is similar to that within population-level risk assessments, Dr. Suter believes. "This is why we do uncertainty analysis," he said. This type of approach has been used to predict extinction threshold effects in populations of fish.

However, given natural environmental variability (e.g., at hazardous waste sites) population-level predictions may be hard to make. Are such predictions too generic to be useful? Would manufacturers have to write labels that reflect the environmental characteristics of every region?

In the Netherlands, scientists sometimes try to describe the relative environmental impact of pesticides. To answer such a question requires an understanding of population-level effects, which they still cannot answer. People are also asking whether pesticides are really more of an environmental threat than other factors. U.S. regulations require EPA to look at the effects of pesticides separately -- and it is something we have some measure of regulatory control over. Indeed, population models may account eventually for a variety of landscape effects.

Generic population -- ecosystem -- level effects can help establish the threshold for mortality and fecundity effects for many chemicals. It is a good scientific base for risk assessments even though it is somewhat imprecise. But at a practical level, individual risk assessments may be easier to refine.

*FIFRA risk assessments.* Under FIFRA, "unreasonable" adverse effects of pesticides must be prevented if at all possible. In the past, the burden of proof rested with industry. Today that responsibility is shifting toward the EPA, which now must also prepare credible risk assessments. Pesticide manufacturers are also conducting their own risk assessments. There is nothing in FIFRA, however, that requires endpoints to be verified in the field.

*Risk management decisions.* Can risk management decisions concerning population effects be made when there are too few data? Should such decisions be made in the risk characterization stage, or should they be considered earlier? Panelists discussed how to handle uncertainty in the context of risk management decisions, with some panelists seeing it as guide and others more worried that the lack of basic data made such predictions extremely unreliable.

According to Dr. Suter, comparative risk assessments are different from those defined by a particular "threshold" effect for an endpoint such as population. If you are comparing old and new pesticides, less absolute predictions are probably acceptable and population-level assessments can be done in a probabilistic manner. Ecosystem-level assessments may not be necessary. Moreover, he asked, "Would a decision based on soil ecosystem process be accepted?"

Dr. Barry did not think it would be difficult to go from dose-response effect to larger scale ecological effects. Population-level stochastic models have been used to devise management strategies for striped bass and PCB's and the models for birds and mammals are similar. This type of "what if" prediction helps refine a risk assessment prediction by providing a best estimate of the consequences.

In comparative risk assessments, the uncertainty is similar for all pesticides. More important is the "distribution" in the uncertainty, including how uncertain you are about an estimate, not whether you are at the 95% confidence level. Dr. Odenkirchen felt that toxicological data for fecundity and mortality are not available, yet Dr. Moore thought it could be obtained (and for high-use pesticides it probably exists). It may be necessary to ask the registrant for additional data.

*Data collection.* The underlying principle in the ECOFRAM report is that you start with basic information and then refine the prediction with as many additional data as required. The discussion focused on the amount of data needed to conduct a probabilistic study (and define threshold effects) and how to use the models to guide data collection efforts.

Dr. Moore feels that uncertainty analysis, even for population-level effects, provides valuable information. Each case should be treated separately and the models used to guide the data collection process. It is important to start with a simple sensitivity analysis. Such an approach is better than point estimates: the more uncertainty you have the more you need uncertainty analysis. The drawback is that may also push us to request more data earlier in the process.

Other panelists felt that there are not enough data to conduct this process as effectively. Dr. Cobb suggested that the report provide more guidance about data collection. Several case studies -- both data rich and data poor -- should be conducted to show whether it is possible to use existing data to make probabilistic predictions at the population level. Dr. Urban advocated using a case study approach to answer question 1A, especially for cases where the endpoints are not birds.

*Other endpoints and exposure models.* We need to consider more than agro-ecosystems and birds when thinking of pesticide effects. Mammals, amphibians, and reptiles represent important endpoints as well. Dr. Moore believes that the form of the exposure models for many species is similar. Dr. Grue recommended that we focus on the most clearly understood groups first -- not worry about all species where data are often lacking -- but that EPA should have the flexibility to look at the ecosystem and decide what data are needed. It is better to focus on the most sensitive species and acquire additional or surrogate data if needed. In other words, apply expert judgment to the process.

Dr. Suter suggested that the kinds of models developed in the human health arena, which are not dependent on lots of expensive toxicity tests but extrapolate in a realistic way between species, be carefully evaluated.

The current exposure model is not very good, and only a very limited amount of such data (primarily related to application of a pesticide) is submitted to the EPA by a registrant. Conditional registrations, with monitoring requirements, are legally possible, but once a pesticide has been used commercially it is very hard to change course. Thus it is important to gather sufficient information about the pesticide's impact during the registration process.

*Alternative tools: expert judgment.* "Expert judgment" can help reduce the uncertainty in situations where there are insufficient data. Since expert judgment is a form of subjective analysis, such methods need to be carefully evaluated. Dr. Cobb opposes using expert opinions in such cases precisely because they are subjective, and would prefer to collect more field data. The ECOFRAM report did not consider these alternative methods because of time limitations.

## **Question 2: Tiers vs. Levels of Refinement**

The discussion focused on the level of refinement and tiers processes for conducting risk assessments -- why flexible tiers are recommended in the aquatic ECOFRAM report and flexible levels of refinement in the terrestrial ECOFRAM report. There was also a discussion about how to account for (reduce) uncertainty in the process; whether standardized data sets and procedures are required in order to have consistent results; and how these probabilistic tools will be used by risk managers to define acceptable risk thresholds.

*Tiers vs. refinement levels.* "Tiers mean boxes," said Dr. Barry, rather than a continuum, which leaves more room for deciding how to proceed and what to study. Tiers can define unwanted decision thresholds, as in the case of cancer rate studies. The ECOFRAM committee used levels rather than tiers because of the desire to have as much flexibility as possible: some parameters can be refined while others do not need to be.

However, there was confusion about the real differences between the two approaches. The overall feeling was that "levels of refinement" was an effective way to utilize a variety of probabilistic risk assessment methods: it enables you to choose the model you need at any point in the risk assessment process and avoid conducting unnecessary exposure and effects tests.

ECOFRAM considers field data a form of refinement if these data are being used to improve the exposure and effects analysis, and risk management decisions generally. Measuring adverse impacts in the field is an altogether different kind of process from that being proposed in the report.

Dr. Suter prefers "tiers" designed around the available information -- rather than model refinement per se -- since data are the main variable. However, he advocates using the most appropriate model which is consistent with the process proposed by the ECOFRAM committee. In general, there will be a refinement in the data used and scenarios considered as you move toward higher level assessments. There is a need to manage uncertainty from the start, even if the initial assessment is largely deterministic.

*Consistency.* There was some discussion about whether the refinement process proposed

by ECOFRAM would lead to inconsistent results. ECOFRAM members suggested that inconsistencies might result from the use of expert judgment, not the flexible framework of options and selection processes they have recommended. The first level, which is more rigid in nature, is likely to be especially useful for making comparisons between pesticides. As you expand the analysis, the process will become more complex and focused.

What do you do in cases where there are not enough exposure data at level 1? According to ECOFRAM, a sensitivity analysis will show if collecting more data will reduce the uncertainty significantly.

Dr. Odenkirchen was especially concerned with the "subjective" nature of the process. Of particular concern is how to guide the selection of effects and exposure tests, and data distributions, in the absence of tiers. ECOFRAM envisions using such subjective data distributions to guide the collection of empirical data and to show what endpoint processes need to be refined. In their view, "what if" predictions would not be used to make the final risk management decision, but would be replaced, or augmented, with more real-world data.

Dr. Barry believes that all parameters, despite their uncertainty, should be left in the models since such uncertainty is an integral part of probabilistic risk assessments. Moreover, expert opinion can be introduced into the assessment to provide a first-order estimate of the risk. Dr. Cobb wants to see more empirical data: lab data or chemical structure data to estimate a half-life, for example.

Dr. Moore pointed out that the distinction between expert opinion and empirical data is not clearcut. Expert opinions are formed partially on the basis of results from previous empirical studies. The design of empirical studies (e.g., where to sample, how many samples) is partially based on expert opinion. Thus, expert opinion and empirical data are not mutually exclusive ways of getting information. Still, Dr. Odenkirchen indicated that EPA would like this process to be somewhat more structured if possible; he would like more guidance concerning the establishment of more "generic" data sets or distributions, and the use of such tools as a sensitivity analysis.

*Data uncertainty.* Probabilistic assessments can incorporate additional subjective knowledge, but the ECOFRAM report did not examine such methods as Bayesian analysis or expert opinion. The value of uncertainty, Dr. Lavine noted, is that it can be used to indicate which assumptions need checking and which assumptions may be scientifically correct. Most panelists thought that this approach can be used successfully by the Office of Pesticide Programs to conduct effective risk assessments, with either generic or more empirical data sets. Historical exposure data sets do exist for birds.

One group clearly favors the use of all existing data, as a first major step in the refinement process, if they are combined with a sensitivity, or bounding, analysis. However, Dr. Odenkirchen noted that EPA has trouble reaching consensus on the value of such parameters as soil half-lives, let alone larger distributions of data. He suggested that it would be better to have some form of standardized, off-the-shelf data sets that could be used by all risk assessors. Moreover -- and

perhaps more importantly -- the exposure and effects parts of the refinement process should receive a similar amount of attention in the analysis. Case studies may help resolve this issue.

*Aquatic vs. terrestrial refinement process.* There was a sense that the refinement processes adopted by the aquatic and terrestrial ECOFRAM committees were not so different: the terrestrial ECOFRAM report tried to identify tools that could be used at a given level; the aquatic ECOFRAM report said less about what tools should be applied. Some of the apparent rigidity in the aquatic tier system is the result of more well defined exposure assumptions. The terrestrial group is at a different stage of development in the risk assessment process.

There seems to be inconsistency in the way the two groups define "endpoints"; this needs to be brought closer together. The terrestrial group has not gone as far as the aquatic group with population-level modeling but envisions using the models in a similar manner. No one wants to end up with the type of risk assessments that are mandated under the Superfund process.

*Risk management decisions (Chapter 6).* As part of a registration application, EPA currently receives a base data set from an applicant, which is then used to develop risk quotients. In cases where the "threshold of acceptability" has been reached or exceeded, EPA may ask for more data. This is a pivotal part of the process, since it is difficult to halt the use of a pesticide once it has reached the marketplace. It is very important that risk managers receive sufficient information early in the registration process so that they can request additional data from the applicant and refine their impact predictions generally.

Currently, there is no attempt to calculate the uncertainty associated with a particular risk assessment. It is not clear yet how such "probabilistic predictions" will affect the actual registration process, but it is clear that they will change the way regulators and manufacturers comply with FIFRA. Hopefully, they will ask for the kind of data needed to develop a more refined risk assessment -- if it will benefit the analysis -- or consider registering the pesticide with additional conditions. The actual benefit of using a particular pesticide must also be carefully weighed (this is generally done in a less quantitative way).

*Refining the assessment: how far?* "The closer the actual risk is to the threshold of acceptability the more precision is required in the risk assessment to enable a decision to be made" (Chapter 6-8). For level 4 assessments, EPA may need to conduct an internal peer review of the process in order to ensure that the decisions are consistent. For the re-registration of a pesticide, outside review might be necessary. How EPA will train its staff to use these new risk assessment methods is a critical issue. Once again, well-developed case studies are likely to be an important training tool.

It is too early in the process to suggest standardized procedures, but EPA needs to consider what endpoints are important and to use the models to help pinpoint the most sensitive species.

### **Question 3: Data Needs for Probabilistic Risk Assessment (Chapters 6 and 7)**



The discussion focus shifted to the adequacy of the existing database: have the data been identified? Can we use existing data, such as single LD50s, to characterize the effect of new chemical species? The less knowledge you have the more you need a probabilistic risk assessment.

*Exposure and effects data.* According to the ECOFRAM committee, Table 6.2.1 summarizes the level of refinement process, not the precise data that will be used in the models. Chapters 3 and 4 address the question more accurately. ECOFRAM has not identified all the data; moreover, they believe that case studies will help show what data are really valuable and what are not.

Short-term exposure data exist for birds but not for most other species. Dr. Urban believes that without case studies it will be difficult to answer this question. New exposure tests will have to be developed to measure longer term dose-response levels, for example. Dr. Lavine suggested that the question makes no sense given the fact that it is a probabilistic risk assessment.

Dr. Suter believes that for acute studies the LD50 is not sufficient; ataxia data (for example) can be gathered during dose-response tests but usually are not; the current reproduction study is not adequate either. As much toxicity and dose-response data as possible should be gathered during reproductive studies. What's needed is more than simply a benchmark value.

*Registration data and extrapolation factors.* ECOFRAM envisioned picking four "generic" species that inhabit an area to characterize the potential ecological risk posed by a particular pesticide. This analysis produces an exposure profile -- or distribution -- that can be used to define the probability of effects for all species of concern. It is hoped that such a profile will depict the range of effects we can expect from a pesticide.

At lower risk refinement levels, sensitivity values for one species are defined using historic data sets. At higher refinement levels, the extrapolation factors used by the ECOFRAM working group are defined by testing the chemical of concern; at lower levels extrapolation factors are based upon tests on other chemicals. The risk assessment methods change above level 2.

Level 1 assessments are not likely to provide enough information about adverse effects. Therefore, it may be more cost-effective for both the applicant and EPA to begin with a higher tier assessment in which several key species are tested. (The uncertainty involved with using one LD50 and the errors associated with the extrapolation factors are likely to be very large.)

Expanding the existing historic data is clearly important, even though the most data-rich pesticides tend to be the most toxic chemicals. Probability analysis exposes the data gaps since it identifies the existing uncertainty in the exposure and effects models. Data gaps can be filled by asking the registrant for additional information or by comparing similar compounds.

The sensitivity of key species needs to be defined. Dr. Lavine is concerned that even 10 species might not reduce the uncertainty inherent in the extrapolation of sensitivity data; even parameter fitting may produce misleading results.

Can generic approaches to describing the sensitivity of a species to a pesticide be developed without a great deal more investment in the collection of new toxicity data?

In some cases new extrapolation factors may have to be developed. Testing species sensitive to some classes of pesticides -- such as raptors, which are more sensitive to certain classes of chemicals -- will provide important toxicological knowledge. It will also diminish the uncertainty by refining the distribution curves.

Dr. Moore suggested that we pay more attention to problem formulation and not mandate additional testing of endpoint species; only the most sensitive species should be tested he believes. Dr. Edwards suggested that such a focus might bias a risk analysis. ECOFRAM wants advice on this issue.

*Higher level refinements and exposure data.* Some panelists felt that the ecological variability associated with a level 4 analysis cannot be quantified. Dr. Suter responded by saying that it is only necessary to consider locally the probability of an adverse effect; it is very important to carefully define the problem that you want to answer and to characterize the dependency among interrelated variables.

Insect residue data may be biased by data collection techniques. When reviewing registration data, it is very important to look at sampling techniques. It should be possible to develop generic "regional" databases for things like precipitation, soil types, and wildlife exposure factors. This should make risk assessments more consistent and easier. A lot of these data are available, even though the original field studies were not designed with probabilistic risk assessments in mind.

Where can we get other data sets for distributions? For example, there is not much information about food preferences of many species of concern. For some species, like kingfishers, which consume a particular size of fish, the uncertainty is minimal; for opportunistic species like many farmland birds it is much larger. Exposure data from stomach analyses may provide additional information about feeding habitats. This variability needs to be expressed in the distributions used in the models. The real ecological behavior of species is important too and can be obtained by careful sampling at higher levels of refinement.

EPA also needs to gain access to the raw data in the Fletcher database; although it does not address such real-world effects as wind drift and multiple application of pesticides. There is nothing equivalent to "Basins" -- a database used by the aquatic risk assessments.

Dr. Grue believes that the real uncertainty is not in the data themselves, but in using such "surrogate" information about one group of species to estimate the exposure levels for other species. Composite sampling techniques can reduce uncertainty in exposure data too. But how many sampling sites are really necessary? EPA must obtain better field data from manufacturers, including environmental factors related to weather patterns.

#### **Question 4: Limitations for Other Endpoints of Concern**

ECOFRAM did not go very far with population models. Dr. Grue noted that "populations" are very difficult to define because of the great spatial heterogeneity in nature. Population models can be probabilistic and are necessary in order to calculate the combined effects of a pesticide. Moreover, the combined threshold effects of a pesticide on two endpoints -- for example, mortality and reproduction -- are much different than if they had been considered independently and must be considered using a population model. Indirect effects, on reproduction success and other endpoints, are not well understood though.

Reproductive studies do not provide the kind of data required in a population model; but you can use the approach applied in human health studies to get reasonable dose-response values. However, reproductive studies may not enable you to predict population-level effects, especially if there is a great deal of uncertainty associated with the population models themselves and the input parameters.

Dr. Grue does not think that everything needs to be based on a probabilistic assessment; existing toxicological data may be sufficient according to Dr. Suter. In addition to basic ecological parameters, it is important to consider how such population-scale information will be used by risk managers. The Patuxent River lab has collected lots of data on bird distributions, including those contained in Gap Analysis and Christmas Bird Count. Such reviews are similar to a hazard assessment.

The registration process is the best time to consider population-level effects; it is difficult to base a registration decision on comparisons between chemicals, even if the new one appears to be much safer than an existing one. This approach may, however, help improve mitigation decisions.

There was a general consensus that the data gaps are real -- and that we must look for as many data as indicated by sensitivity analysis.

## Question 5: Research and Validation

*Case studies: validation of models?* ECOFRAM does not believe that models can be field validated directly but that some form of verification of the process is possible. Some understanding of the uncertainty in the data that are used to run the models can be obtained.

Validation of probabilistic models is very difficult -- only a crude validation, or verification, of models may actually be feasible. It may be a bit easier to validate the risk assessment process generally. Dr. Odenkirchen would like to compare model exposure concentrations with chemical concentrations determined by field methods. Dr. Barry believes that model calibration is very site specific.

Risk managers should be involved in the implementation process, as should some members of the ECOFRAM working group. The three main priorities are:

- \* Conduct several case studies, perhaps within EPA.
- \* Establish a working group composed of people from industry, EPA, and academia who would direct the case studies and research plans.
- \* Evaluate the policy implications of these new tools, including how to integrate the cost-benefit analysis, other decision tools, and probability-based predictions into risk management decisions.

Also, encourage greater industry and regulatory collaboration, perhaps using a CRADA-like structure as the model; access proprietary data. However, it may take some time to see what data are the most useful. Establish a generic list of species that everyone feels are the most sensitive species. Unfortunately these ideas were not implemented once before, because such data gathering efforts were not considered essential.

The current models have not been field tested either. We need to show that what we think is happening is indeed occurring in nature, especially for the most commonly used pesticides. Understanding why models do not work is usually very hard to ascertain. Filling the data gaps identified by sensitivity analysis will be an essential part of this process.

## Question 6: Approaches to Case Studies

Case studies will help show how the risk assessment and management approach works and how specific decisions are made throughout the process, and expose some of the weaknesses in generic exposure and effects data. Case studies were a high priority of the aquatic workshop panel too.

There was a general consensus that the best way to proceed is to simulate several risk assessment situations: one a registration process where there are lots of data and another where there are few available data. It is important to test different scenarios, including situations where a deterministic-based model was used to make a risk management decision.

One approach is to let the agency implement the case studies, using a multistakeholder coordinating group. Input from both industry and academia is essential. Alternatively, several groups could work on a case study separately, or a neutral group could be hired.

Risk managers should be involved in case study as well. It may be better not to identify the study compound (to avoid any bias), or substitute a simulated compound for a real one. It is important to find out how different groups make decisions. It is also important to see what additional information risk assessors and managers request from pesticide manufacturers. Some people may be very conservative; others will not be comfortable with lots of probability. Dr. Barry suggested that model results may not be very consistent or accurate even when sufficient data are available (as for example after "Chernobyl") because a great deal of expert judgment is involved in the process.

Each group would be required to document how they reached key decisions, and they would also have to define the problem. ECOFRAM considered five case studies, but to the problem formulation stage only.

Some panelists would like to see the process become more international in character, in order to take advantage of the different approaches being developed by OECD countries. Indeed, there is a current European proposal to conduct several case studies that would be similar in scope to what is being proposed here.

Initial case studies should be relatively simple. They need not be concerned with developing generic databases on many species, but with establishing a well-defined "decision tree." At higher refinement levels, indirect effects can be introduced into such case studies. The panel felt that case studies are more of a training exercise -- a way to better understand the probabilistic risk assessment process generally.

## **PUBLIC COMMENT**

### **Tom Bailey, Office of Pesticide Programs, U.S. EPA**

Mr. Bailey emphasized the importance of establishing a balance between risk assessment and risk management. Risk managers, he said, must understand this new process in order for it to be successfully implemented. Moreover, he believes that it is incumbent upon risk assessors to help risk managers learn this process and to help them design a framework for its effective use.

And, although the proposed risk assessment process is clearly more complex, he believes that it will still be necessary to register pesticides in a timely fashion.